

Reinforcement Learning and Optimal Control

A Selective Overview

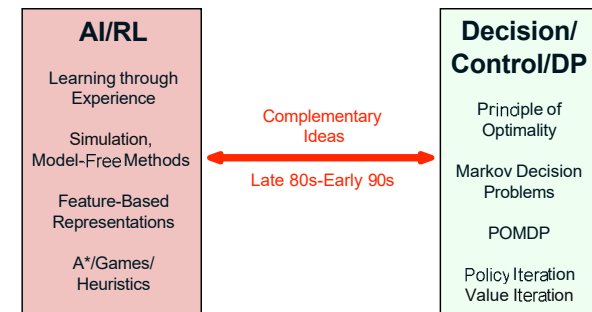
Dimitri P. Bertsekas

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology

2018 CDC

December 2018

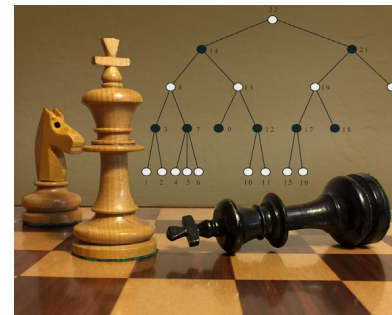
Reinforcement Learning (RL): A Happy Union of AI and Decision/Control Ideas



Historical highlights

- Exact DP, optimal control (Bellman, Shannon, 1950s ...)
- First major successes: Backgammon programs (Tesauro, 1992, 1996)
- Algorithmic progress, analysis, applications, first books (mid 90s ...)
- Machine Learning, BIG Data, Robotics, Deep Neural Networks (mid 2000s ...)
- AlphaGo and Alphazero (DeepMind, 2016, 2017)

AlphaGo (2016) and AlphaZero (2017)



AlphaZero

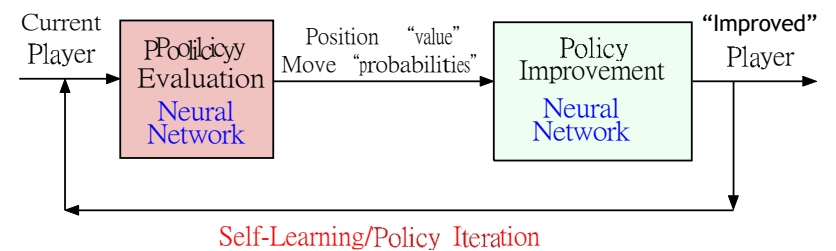
Plays much better than all chess programs

Plays different!

Learned from scratch ... with 4 hours of training!

Same algorithm learned multiple games (Go, Shogi)

AlphaZero was Trained Using Self-Generated Data



- The "current" player plays games that are used to "train" an "improved" player
- At a given position, the "move probabilities" and the "value" of a position are approximated by a deep neural net (NN)
- Successive NNs are trained using self-generated data and a form of regression
- A form of randomized policy improvement **Monte-Carlo Tree Search** (MCTS) generates move probabilities
- AlphaZero bears similarity to earlier works, e.g., **TD-Gammon** (Tesauro, 1992), but is more complicated because of the MCTS and the deep NN
- The success of AlphaZero is due to a skillful implementation/integration of known ideas, and awesome computational power

Approximate DP/RL Methodology is now Ambitious and Universal

Exact DP applies (in principle) to a very broad range of optimization problems

- Deterministic $\longleftrightarrow$ Stochastic
- Combinatorial optimization $\longleftrightarrow$ Optimal control w/ infinite state/control spaces
- One decision maker $\longleftrightarrow$ Two player games
- ... BUT is plagued by the **curse of dimensionality** and **need for a math model**

Approximate DP/RL overcomes the difficulties of exact DP by:

- **Approximation** (use neural nets and other architectures to reduce dimension)
- **Simulation** (use a computer model in place of a math model)

State of the art:

- **Broadly applicable methodology**: Can address broad range of challenging problems. Deterministic-stochastic-dynamic, discrete-continuous, games, etc
- There are **no methods that are guaranteed to work** for all or even most problems
- There are **enough methods to try with a reasonable chance of success** for most types of optimization problems
- **Role of the theory**: Guide the art, delineate the sound ideas

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/625042030034011034>